数理解析研究所講究録
179巻 1973年 62-73

文脈自由形言語の情報論的性質

NHK総合技術研究所　尾関　和彦

§1. まえがき

自然言語であれ人工言語であれ，言語の持つ最も重要な機能のひとつは情報の伝達であろう．それでは言語が伝達する情報量というものをどのように考えたらよいであろうか．

情報理論が創始されて以来，自然言語に対してもMarkov過程のモデルが考えられ，それにもとづいて一文字当りの平均情報量が計算されたこともある．Markov過程のモデルを考えることは，その言語の文法として正規文法を考えることにほかならないが，近年の言語理論の進歩は正規文法が自然言語の文法として不十分であることを明らかにした．ここでは文法のクラスとして，自然言語の構造の基本的な部分を記述することができるといわれている文脈自由形文法をとりあげ，それを言語能力として持つ話し手のモデルとして，プロダクションに確率が付与された，いわゆる確率的文脈自由形

文法を考える．そしてその話し手の話す言語について情報論的議論を展開する．

§2．言語が情報をになう機構

ここに二人の人A，Bがいるとし，彼らは共通の言語を話すとする．彼らの言語能力のモデルとしてつぎのような文脈自由形文法Gを仮定しよう：

$G=(V_T, V_N, P, S)$，

$V_T=\{$花，鳥，娘，は，とても，美しい$\}$，

$V_N=\{S$，主部，述部，名詞$\}$，

$P=\{S\rightarrow$主部述部，主部$\rightarrow$名詞は，名詞$\rightarrow$花，名詞$\rightarrow$鳥，名詞$\rightarrow$娘，述部$\rightarrow$美しい，述部$\rightarrow$とても述部$\}$．

AおよびBが話しかつ理解することのできる文の全体，すなわち彼らの言語は，$\{$花は美しい，娘は美しい，---，花はとても美しい，---$\}$という可算無限集合である．いまAがBに

‘娘はとても美しい’ (1)

と言ったとしよう．この文がAの伝えたい情報をにない，かつBがこの文を受けとってAの意志を知ることができる機構はどのようなものであろうか．(1)の生成過程を考えてみる

と，

S ⇒主部述部⇒名詞は述部⇒娘は述部⇒娘はとても述部
⇒娘はとても美しい　(2)

というステップを踏んで(1)が生成されることがわかる．この生成過程の第3ステップにおいて名詞→娘というプロダクションが適用されているが，Aはここで名詞→花，あるいは名詞→鳥というプロダクションを選ぶこともできたはずである．つまり三つの可能なプロダクションの中からAは特に，名詞→娘というプロダクションを選択したわけである．このようにして(1)は‘花’でもない‘鳥’でもない‘娘’は…というAの意志をになうことができるのである．一方Bは(1)を受けとり，その生成過程の第3ステップにおいて三つの可能なプロダクションの中から特に名詞→娘が選択され適用されていることを見出しAの意志を知ることができるであろう．すなわち文が情報をになうことができるのは，その生成過程の各ステップにおいて適用するプロダクションを話し手が選択できるからであり，聞き手が文を受けとって情報を得ることができるのは，その文の生成過程の各ステップにおいていくつかの可能なプロダクションの中から実際にどのひとつが選択されたかを知るからにほかならない．このように文を生成過程に分析して考えれば‘情報を得るとはいくつかの可能

性の中からどれが選択されたかを知ることである，という情報に対するShannonの基本的な考え方がそのまま適用できることがわかる．

ここでは話し手ー聞き手の言語能力のモデルとして文脈自由形文法を考え，話し手の意志を，各プロダクションにそれが適用される確率を付与するという形でモデル化する．つまり確率的文脈自由形文法を話し手のモデルとする．それをもとにして文脈自由形言語に情報論的考察を加えていこう．

§3. 確率的文脈自由形言語

記法1. 文脈自由形文法（cfg）Gを通常の記法にしたがい$G=(V_T, V_N, P, S)$と表わす．ここでV_Tは末端記号の集合，V_Nは非末端記号の集合，Pはプロダクションの集合，$S\in V_N$は出発記号である．

定義1. 確率的文脈自由形文法（scfg）G_sは順序対(G, φ)である．ここで$G=(V_T, V_N, P, S)$は既約なcfgであり，φはPから実数区間$(0, 1]$への写像である．ただし$\varphi(P_i)$，$P_i\in P$を同一の左辺を持つP_iについて加えたものは1に等しいとする．GをG_sの台という．

<u>記法2.</u> $G=(V_T, V_N, P, S)$ を cfg とするとき $\alpha\in(V_T\cup V_N)^*-V_T^*$ から出発する G による左生成過程の全体を D_α, D_α の元で文生成過程となっているものの全体を E_α と書く. D_α, E_α の元で n ステップのものの全体をそれぞれ $D_\alpha^{(n)}$, $E_\alpha^{(n)}$ と書く. 特に D_S, E_S, $D_S^{(n)}$, $E_S^{(n)}$ をそれぞれ D, E, $D^{(n)}$, $E^{(n)}$ と書く.

<u>記法3.</u> $d\in D_\alpha^{(n)}$ の第 i ステップ $(i=1,\cdots,n)$ においてプロダクション P_i が適用されているとき, $d=P_1\cdots P_n$ と書く.

<u>定義2.</u> $G_s=(G,\varphi)$ を scfg とする. そのとき写像 ψ_α: $D_\alpha\to(0,1]$ をつぎのように定義する: $d=P_1\cdots P_n\in D_\alpha$ に対して $\psi_\alpha(d)=\varphi(P_1)\cdots\varphi(P_n)$.

<u>定義3.</u> 写像 $\psi'_\alpha: 2^{D_\alpha}\to[0,\infty)$ をつぎのように定義する; $M\subset D_\alpha$ に対して $\psi'_\alpha(M)=\sum_{d\in M}\psi_\alpha(d)$, ただし $\psi'_\alpha(\phi)=0$.

ψ'_α は ψ_α の拡張であるが, 両者をともに ψ_α で表わす. ψ_S を ψ と書く. ψ の定義域を E に制限したものも ψ で表わす.

<u>命題1.</u> $(E, 2^E, \psi)$ は測度空間である.

<u>定義4.</u> Eの元に，それによって生成される文を対応させる写像をhとする．

<u>記法4.</u> $h(E)$ を $L(G)$ と書く．

hによってEから$L(G)$上にひきおこされる測度を考えることができる．すなわち，

<u>命題2.</u> $\tilde{\Psi} : 2^{L(G)} \to [0, \infty)$ を $M \subset L(G)$ に対して $\tilde{\Psi}(M) = \Psi(h^{-1}(M))$ と定義すると $(L(G), 2^{L(G)}, \tilde{\Psi})$ は測度空間である．

$(L(G), 2^{L(G)}, \tilde{\Psi})$ を G_s によって生成される確率的文脈自由形言語（scfl）という．

<u>命題3.</u> $\tilde{\Psi}(L(G)) = \Psi(E)$．

<u>命題4.</u> $G_s = (G, \varphi)$ を scfg とする．Gがあいまいでなければ $(E, 2^E, \Psi)$ と $(L(G), 2^{L(G)}, \tilde{\Psi})$ は測度空間として同型である．

定理1. 任意のscfgによって生成されるscfl $(L(G), 2^{L(G)}, \tilde{\Psi})$ について $0 < \tilde{\Psi}(L(G)) \leqq 1$ が成立する.

定理1を証明するために二つの補題を用意する.

補題1. 任意のscfgに対して $\Psi(D^{(n)} - E^{(n)}) = \Psi(D^{(n+1)})$ がなりたつ.

補題2. 任意のscfgに対して $\Psi(D^{(n)} \cup \bigcup_{i=1}^{n-1} E^{(i)}) = 1$.

(定理1の証明) 命題3により $\tilde{\Psi}(L(G)) = \Psi(E)$ であるから $0 < \Psi(E) \leqq 1$ を証明する. $E = \lim_{n\to\infty} \bigcup_{i=1}^{n} E^{(i)}$ であり, また $\{\bigcup_{i=1}^{n} E^{(i)}\}$ $(n=1, 2, \cdots)$ は単調非減少な集合列であるから測度の一般論により

$$\Psi(E) = \Psi(\lim_{n\to\infty} \bigcup_{i=1}^{n} E^{(i)}) = \lim_{n\to\infty} \Psi(\bigcup_{i=1}^{n} E^{(i)}).$$

$M_1 \subset M_2 \subset E$ ならば $\Psi(M_1) \leqq \Psi(M_2)$ だから $\{\Psi(\bigcup_{i=1}^{n} E^{(i)})\}$ $(n=1, 2, \cdots)$ は単調非減少な実数列であり, また補題2により $\Psi(\bigcup_{i=1}^{n} E^{(i)}) = 1 - \Psi(D^{(n+1)}) \leqq 1$ であるから $\Psi(\bigcup_{i=1}^{n} E^{(i)})$ $(n=1, 2, \cdots)$ は有界である. したがって $\lim_{n\to\infty} \Psi(\bigcup_{i=1}^{n} E^{(i)})$ が存在する. この値を l とすれば, $0 < \Psi(\bigcup_{i=1}^{n} E^{(i)}) \leqq 1$ であ

るから $0 < l \leqq 1$.　　　　（証明終）.

<u>例1.</u> $G = (\{a, b\}, \{S\}, \{S \to a, S \to bSS\}, S)$ を台として持つ scfg $G_s = (G, \varphi)$, $\varphi(S \to a) = p_1$, $\varphi(S \to bSS) = p_2$, $p_1 + p_2 = 1$, $p_1 > 0$, $p_2 > 0$ を考える．この G_s によって生成される scfl について $\tilde{\Psi}(L(G))$ をもとめると

$$\tilde{\Psi}(L(G)) = \begin{cases} 1; & p_1 \geqq \frac{1}{2} \text{ のとき}, \\ \dfrac{p_1}{1-p_1}; & \frac{1}{2} > p_1 > 0 \text{ のとき}. \end{cases}$$

例1は scfl が確率空間になる場合とそうでない場合のあることを示しているが，scfl が確率空間になるということの意味についてもう少し考察を加えてみよう．いま scfg を文を生成するオートマトンと考え，このオートマトンは文が生成されたときに停止するものとする．$D^{(n)} - E^{(n)}$ は n ステップの生成過程のうち文生成過程でないものの全体であり，$\bigcup_{i=1}^{n} E^{(i)}$ は n ステップ以下の文生成過程の全体である．したがって $\Psi(D^{(n)} - E^{(n)})$ はこのオートマトンが n ステップ以下では停止しない確率，$\Psi(\bigcup_{i=1}^{n} E^{(i)})$ は n ステップ以下で停止する確率である．さてすでに示されたように

$$\tilde{\Psi}(L(G)) = \lim_{n \to \infty} \Psi\left(\bigcup_{i=1}^{n} E^{(i)}\right)$$

であり，$\Psi(\bigcup_{i=1}^{n} E^{(i)})$ は n に関して単調非減少であるから，'$\tilde{\Psi}(L(G))=1$' と '任意の $\varepsilon>0$ に対して自然数 n_0 が存在して $\Psi(\bigcup_{i=1}^{n_0} E^{(i)})>1-\varepsilon$' とは同値である．すなわち，つぎの命題が得られた．

命題5．　つぎの (i), (ii) は同値である．

(i) scfg G_s によって生成される scfl が確率空間となる．

(ii) 任意の $\varepsilon>0$ に対して自然数 n_0 が存在し，オートマトンとしての G_s が n_0 ステップ以下で停止する確率が $1-\varepsilon$ より大きい．

命題5の (ii) を直感的にいえば 'G_s が殆んど確実に有限ステップで停止する' となる．話し手のモデルとしてはそのような G_s のみを考えておけば十分であろうから，以下の議論では scfl はつねに確率空間であるとする．

§4. 言語の情報量

4.1 文生成過程および言語のエントロピー

定義5．　scfg G_s による文生成過程の全体 E のエントロピー H(E) を

$$H(E) = -\sum_{d \in E} \psi(d) \log_2 \psi(d)$$

と定義する．

H(E) は文生成過程ひとつ当りの平均情報量，すなわちP.マーカーを一つ受けとったときに得られる情報量の期待値である．

定義6. $G_s = (G, \varphi)$ をscfgとする．G_s によって生成されるscflのエントロピーH(L(G))を

$$H(L(G)) = -\sum_{w \in L(G)} \tilde{\varphi}(w) \log_2 \tilde{\varphi}(w)$$

と定義する．

H(L(G)) は G_s によって生成される文ひとつ当りの平均情報量である．

あとでわかるように一般に $H(L(G)) \leqq H(E)$ であるから $H(E)$ が有限ならば $H(L(G))$ も有限である．しかし $H(E)$ は常に有限とは限らない．P.マーカーをひとつ受けとったときに，それから無限の情報を得ることが期待できるということは一見不合理のようであるが，以下に述べることがこの疑問を解決するであろう．

定義7. $N: E \to R$ をつぎのように定義する：$d \in E$ に対して $d \in E^{(n)}$ のとき $N(d) = n$.

N は $(E, 2^E, \psi)$ 上の確率変数であり，$N = n$ となる確率 $\psi\{N = n\}$ は $\psi(E^{(n)})$ で与えられる．N の平均値を $M(N)$ と書く：$M(N) = \int_E N\, d\psi = \sum_{n=1}^{\infty} n\, \psi(E^{(n)})$.

定理2. $H(E)$ は $M(N)$ が有限なとき，かつそのときに限り有限である．

$M(N)$ は文生成過程の平均ステップ数であるから，G_s を話し手のモデルと考える以上 $M(N)$，したがって $H(E)$ は有限と考えてよいであろう．

4.2 言語のあいまいさ

言語にはあいまいさのある場合がある．これは同一の文を生成する複数個の生成過程が存在する場合で，文が与えられてもそれがどの生成過程によるものか聞き手が決定できず，あいまいさが残るわけである．

ある scfg によって生成される二つの測度空間 $(E, 2^E, \psi)$, $(L(G), 2^{L(G)}, \bar{\psi})$ について考える．$w \in L(G)$ が生成されたと

いう条件のもとでの $d \in E$ の確率 $P_w(d)$ は条件付確率の定義により $P_w(d) = \psi(\{d\} \cap h^{-1}(w)) / \psi(h^{-1}(w))$ で与えられる.

定義8. $w \in L(G)$ のあいまい度 $A(w)$ を

$$A(w) = -\sum_{d \in E} P_w(d) \log_2 (P_w(d))$$

で，また $L(G)$ の平均あいまい度 $A(L(G))$ を

$$A(L(G)) = \sum_{w \in L(G)} \tilde{\psi}(w) A(w)$$

で定義する.

命題6. $A(L(G)) = H(E) - H(L(G))$.

命題7. $A(L(G)) = 0$ となるのは G があいまいでないとき，かつそのときに限る.

謝辞　この研究を進めるに際しては，NHK総合技術研究所内　数理工学研究会の諸氏に討論をしていただいた．とくに例1の結果は上坂吉則氏および坂井徹男氏に負うところが大きいことを記して謝意を表したい.